

The identification of persons in the middle ages: results from the first "Freiburger Bürgerbuch" (1341-1416)

Portmann, Urs

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Portmann, U. (1981). The identification of persons in the middle ages: results from the first "Freiburger Bürgerbuch" (1341-1416). *Historical Social Research*, 6(2), 11-26. <https://doi.org/10.12759/hsr.6.1981.2.11-26>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

THE IDENTIFICATION OF PERSONS IN THE MIDDLE AGES:
RESULTS FROM THE FIRST "FREIBURGER BÜRGERBUCH" (1341-1416)

Urs Portmann⁺

A semi-automized code has been used to identify individuals who are referred to on more than one occasion in the source, the "Freiburger Bürgerbuch" (1341-1416). This code is made up of parts of the family and first names. It allows the orthographic and phonetic variants of the same name to be taken into account. The plausibility of the grouped names is then automatically tested. The final decision about the identity of the names has, however, to be taken by the historian.

The identification of persons, property and events is a well known procedure in historical research. Every compilation and representation of data concerning persons, groups of persons, events or property purports to add together information concerning identical units. For a long time this identification process was based only on the historical experience of the inquirer, but the use of electronic data processing today obliges the historian to explain his methods and thus the whole process of identification becomes more accurate and easier to control. (1)

In the following paper, we shall describe the method used to identify persons in the project "Freiburger Bürgerbuch 1341-1416". (2) The first "Bürgerbuch" of the town of Freiburg in Switzerland from the year 1341 to 1416, is the most important medieval source of this kind in Switzerland. This source contains 2200 inscriptions with 7000 names. (3) Each inscription includes the name of the "Bürger" his occupation, social position, place of origin and kinship relations. It also includes a "security-object", such as a house, which can be situated from its road name, its owner and from the owners of the two neighbouring houses. The same person can therefore appear frequently in the book: as a "Bürger", as an owner, or as a relative of someone else.

The first stage in the systematic investigation of this source is the identification of all the different inscriptions of the same person. The identification of property based on a person's identity will take place in a second stage. Although we encountered the same problem of insufficient information concerning the persons and their changing names as in most medieval sources, this identification process of coordinating the names is essential for a historical investigation of this list.

⁺ Address all communications to: Urs Portmann
Séminaire d'histoire médiévale,
Rue Cité-Devant 2
CH - 1006 Lausanne

In the first part of the paper we describe the method employed in this project and in the second part we give a critical appreciation of the above method.

1. The method employed

In spite of the lack of information available in this source and the restricted use of electronic data processing in our project, we used the same concepts of comparison space, comparison function and decision rule in the process of record linkage or identification as are used in modern historical research. (4)

In our case, the comparison space is composed of a stock which includes all the data from the "Bürgerbuch" in the form of fixed format name-records. Each name-record has its own identification-code which is created automatically by the machine on the basis of the first and family name. This code must exclude all phonetic and transcriptional variations.

The comparison function is divided into two stages: the searching step and the comparison step. During the searching step the name-records are sorted using an identification-code and then grouped together. In this way all records with the same possible names are in one "pocket". This identification-code is called a "pocket variable". (5) At the comparison stage, the historian has to check all the name-records of each group, as the computer only shows the most relevant information concerning logical oppositions. We can therefore call this second stage a semiautomatic process.

The decision rule designates two name-records as linked, non-linked or as indeterminate. We shall show that the lack of information in all medieval sources authorizes simple rules of decision.

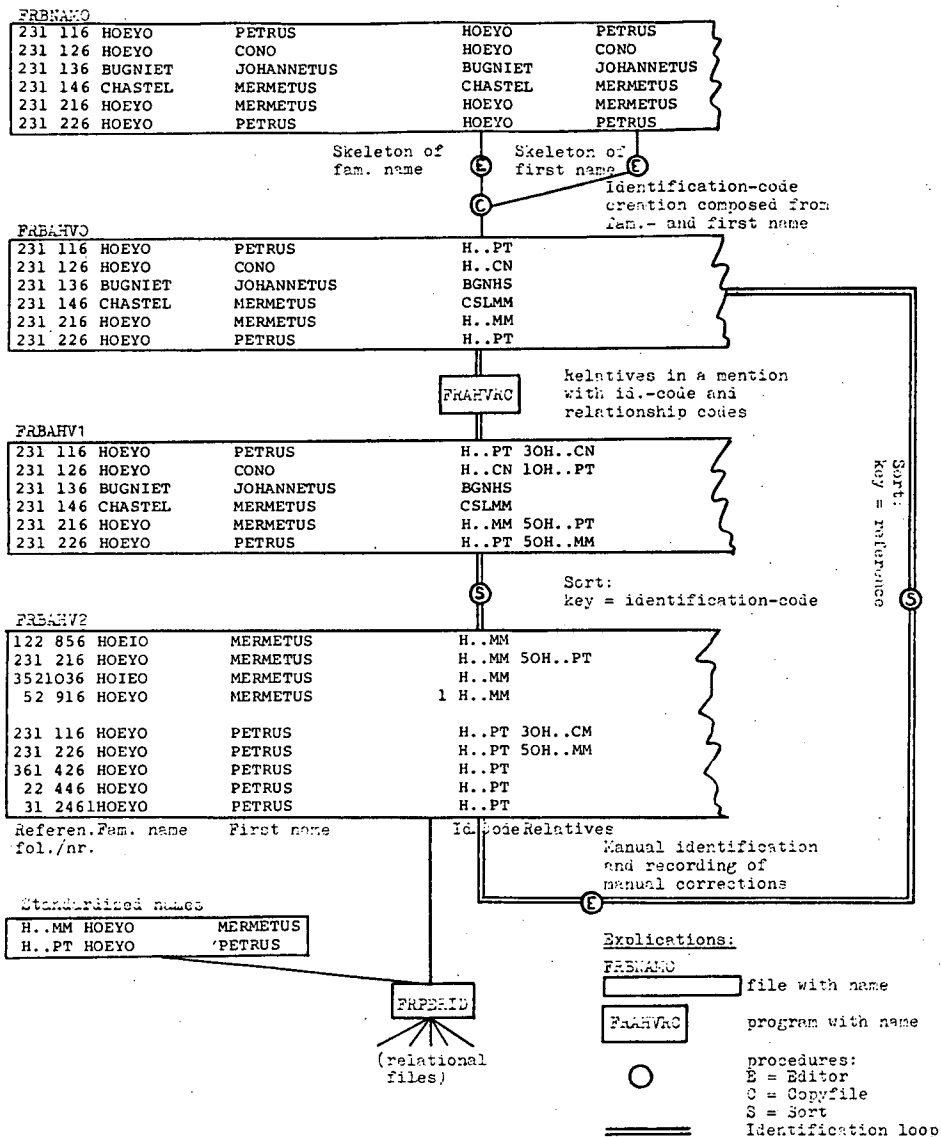
A program creates a personal record with all the name-records that concern the same person, and which includes all the information relating to that person.

The identification-code

The most important and complete information for the identification process is the first and the family name. Any other information is too scanty to add to the identification-code. Although the first and family names exist in a complete form, they present certain difficulties. The names are given in Latin, German or French. They are sometimes abbreviated or transcribed badly. Names can change: the occupation or place of origin can become the family name. One comes across homonymic names: different people have the same name. In spite of all these problems, the historian must build up a full identity based on the name. Therefore it is important to keep the names as written in the source on the name-record so that every identification can be easily controlled. (6) Then all the first and family names are re-copied a second time on the name-record so that all modifications are only carried out on this copy. (fig. 1, file FRBNAMO)

Fig. 1

Short identification processing run



The identification-code must embrace many spelling and idiomatic variations of the same name and at the same time must try to be as restrictive as possible. We try to reach this target using a skeleton formed from the first name and the family name. These skeletons are built up in a pragmatic way without paying any attention to philological rules. First of all, all double consonants are reduced and vowels are suppressed. Then determined groups of characters and entire names are considered to be the same. The skeletons are created by an Editor-procedure (7) that is easy to establish and can be added to if necessary. After the command CHANGE (reduced to C) follow the characters or a group of characters which must be changed. These in turn are followed by the required form; for example: command "C/BIGY/BYE/*" will change all names "BIGY" into "BYE". The first name and the family name are treated separately. The rules used for the treatment of the family name are the following:

1. Entire names are standardized:
C/BIGY/BYE*
108 commands
2. Groups of consonants are standardized:
C/CQ/Q/*
12 commands
3. Double consonants are reduced:
C/TT/T/*
9 commands
4. Last s and z are eliminated:
C/S //*
2 commands
5. All vowels except the initial vowel are eliminated:
C/A//**
6 commands
6. Different groups of consonants are standardized:
C/SVL/CVL/*
4 commands

The processing concerning the first name is similar and involves 56 commands. (8)

The identification-code is created from the first three characters of the family name skeleton and the first two characters of the first-name skeleton. A sixth place remains free in order to distinguish manually identical codes concerning different persons.

name:	KUBLERS	first name:	HENSLI
	KUEBLER		HENSILLINUS
skeleton:	KBLR	skeleton:	HSL
	KBLR		HSLNS
identifica- tion code:			KBLHS

The base-list for manual identification contains this code and all other necessary information from the source needed to identify the person.

The comparison function

The comparison function, the main identification process, is carried

out in two steps. In the searching step, all records which probably belong to the same person are grouped together. In the comparison step the grouped records are checked to find out whether they are homogeneous, and separated if necessary. (9) Addition and separation are described individually in this paper but in practice they take place at the same time in the manual identification process.

The sorting and grouping process is based on the identification-code. In this code, all different spelling and idiomatic variations should have been grouped together. (fig. 1, file FRBAHV2) All name-records which belong together should be contained in one group. (10) Our experience shows us that 12% of all the name-records are not correctly grouped. (11) Two thirds of the 12% are spelling variants that cannot be taken into consideration by the skeleton processing. The other third concerns name modifications. Modifications of a name can be checked through other lists which have been prepared, such as the first name list, occupation, place of origin or nickname list. It is thus possible to identify the name-record of a person whose name has changed or been misspelt, and who has therefore been incorrectly grouped by the identification-code. The first name list is particularly important in this case as the first is more constant and more recognizable for a personal description than the family name. The identification-code in the falsely grouped records is changed manually. After these modifications, all records which concern the same person should be grouped together. Now the historian has to check whether all the records in each group concern the same person.

This check of the name-records in a group is called the comparison step. Unlike automatic comparison methods used in modern historical research, our comparison step remains purely manual in this project. The comparison step and decision rules are taken together in our identification process. The most important identifying characteristics are the family name and the first name. If the two names are identical and there are no other opposing characteristics, we can suppose that they belong to the same person. In fact it is very difficult to distinguish two similar names and their real identities. To illustrate the problem: Are the first names "Hensillinus" and "Hansonus" two different names, or is this a single name written two different ways? Sometimes we find a solution using a supplementary identifying item. A good personal knowledge of the name material is essential in this work and this is the reason why the comparison step would take too long with automatic processing. As the supplementary identifying items are often insufficient, we must take into consideration that many identical names which in fact describe different persons cannot be distinguished from one another. However the main result is not essentially falsified. One could naturally change the processing and group together only the name-records with supplementary identifying factors, but in this case only a third of the records could be identified. All other information would be lost. So we decided to take into consideration all possibly identical records and to renounce measuring the degree of relationship between these identical name-records.

The supplementary identifying items give information about: (12)

- naturalisation: a person cannot normally be naturalized twice.
- place of origin (7%)
- occupation (22%)

- kinship relations: one reference per name 22%
- two references per name 4%
- three or more ref. per name 1%
- total 27%
- dates of known life span
- sex: as the sex is determined by the first name, this is only an indirect identifying item.

The separation of name-records with the same code is done by adding a number in the sixth position of the identification-code. At first corrections are made manually on the printed list (fig.1, file FRBAHV2) and afterwards they are introduced into the computer using a terminal. (fig.2)

fig.2

Name-records division with identical identification-codes

FRBAHV2 (old file containing the corrections)

1332 426	GODION	JOHANNODUS	GDNHS 10GDNHS	7
1332 416	GODION*	JOHANNODUS	GDNHS130GDNHS	/
952 836	GREDE	HANSO	GDNHS2	/
322 626	GREDE	HANSONUS	GDNHS2	/
9221046	GREDE	JOHANNES	GDNHS2	/
9301216	GREDE	JOHANNES	GDNHS2	/

FRBAHV2 (new corrected file)

1332 426	GODION	JOHANNODUS	GDNHS 10GDNHS1	7
1332 416	GODION*	JOHANNODUS	GDNHS130GDNHS	/
9221046	GREDE	JOHANNES	GDNHS2	/
9301216	GREDE	JOHANNES	GDNHS2	/
952 836	GREDE	HANSO	GDNHS2	/
322 626	GREDE	HANSONUS	GDNHS2	/

The two first names design father and son. The four following names go into the same group characterised by the identification-codes.

However, as only some of the name-records with identical family and first names can be differentiated with the help of supplementary information, the others, with their single comparison items, have a question mark written on the identification-code of their name-record

and are considered to be non-identified.

Continual identification

This first stage of processing must be improved by numerous other tests. A program, based on the supplementary identifying items already mentioned, tests the name-records of each person to see whether there are any logical contradictions. The data base also allows for other tests later on, to show up any further contradictions in the records.

Extra information from other sources can change the identification of a person. This change must be registered manually on the identification-code. Once this code is modified, an automatic run will in turn make all the necessary changes in the file (e.g. kinship relations). This identification run is shown in figure 1. It works as follows: 1. The data file is sorted using the reference of the initial source. (file FRBAHVO) (13) 2. The program FRAHVC automatically attributes the corrected identification-code to the family relations of the person. (file FRBAHV1)(14) 3. This file is sorted and grouped according to this new identification-code (file FRBAHV2) and it is with this corrected file that we make any further identifications. Corrections can be continually introduced into the FRBAHV2 file and this file will be sorted once again using the reference of the initial source. Then the program FRAHVC will attribute these new codes to all the relations of the family, this file will be sorted and then the FRPERID program will create a new personal file with the correction. Thus, after the identification-code has been changed manually, all other necessary corrections are carried out automatically except for the standardized name-list.

Standardized names and the creation of a personal record

All the information concerning an individual is collected in the personal record. This record must not contain any name variants, only a standardized name. The program FRPERID creates a list of the first form of the name encountered for each person. Every person must be mentioned on this list, with his family name, first name and nickname. However, these names are still recorded in the written form of the source and they must be standardized with the collaboration of philological specialists. The rules of standardization are as follows: (15)

- Local place names are changed into the modern form of the city/place concerned.
- Names still in use today are modified to their modern spelling.
- All other names are adapted to modern spelling rules.

The FRPERID program creates a personal record from the name-records with the same identification-code and corresponding standardized name. This record contains all the information concerning the person, his occupation, social position, father and place of origin. From the dates found in the source we can calculate the years of a part of the life of the person: first reference, last reference and first reference as dead. This person is described by his standardized name. The

FRPERID program also creates other files, known as relational files, for a relational data base. Thus all indications about naturalization and all changes concerning it will be in the relational file: person-naturalization. Another file is created with the original name-record with all the written variations. Another file contains the kinship relations (see fig. 3), and another political functions. Cross-references to all the persons involved are made using the identification-code. The relational disposition of the information provides two major advantages: firstly, it shows all logical and clearly defined relationships between different given facts, and secondly, any eventual use made of these facts has no influence on the internal organisation of the information.

fig.3

Person-relationship file

(At beginning, example as in fig.1)

H..MM 50H..PT	1. Identification-code
H..PT 30H..CM	2. Relationship
H..PT 50H..MM	(f.i. 10 = father
HF... 60HF.CR	30 = son
HF.CR 60HF...	50 = brother
HGNHB 60HGNJQ	60 = husband)
HGNJQ 65HGNHB	3. Identification-code
HLMCN 10HLMRC	of referred person
HLMRC 30HLMCN	
HLTHS130HLTWL	
HLTHS130HLTSL	
HLTSL 10HLTHS1	
HLTSL 65HLTWL	
HLTWL 10HLTHS1	
HLTWL 60HLTSL	

1. 2.3.

The introduction of supplementary name-records later, from other sources, should be facilitated by this identification method. In this case, firstly, the existing skeleton program generates the identification-codes; secondly, the new name-records are linked with the old file with the help of these identification-codes, but only the historical inquirer decides whether the identification is valid. The FRPERID program is capable of creating a new personal file and other different relational files. If the new data contains a new kind of information, for example about taxes, we have to create a new relational file. As the different files are automatically created, this is an advantage when introducing new sources. Better skeleton programs make it easier to introduce additional new sources.

2. A critical examination of the above method

The identification method for the name-records has been described, but now it must be fully explained and criticized in the light of our experience and our results.

The identification-code serves to suppress all different spelling and idiomatic variations of the names. But does this skeleton processing really group all possibly-identical names together? A test with all the name-records is not possible in our project because the researcher can seldom be completely certain that two names really go together. There is only one exception to this rule. In the year 1416, all living "Bürgers" were transcribed from the old book into a new one with a reference concerning the previous place in the old book. This information gives us a sure way of checking the correct attribution of names. However, although one expects these names, which represent 7% of all the name-records in the project (= 568 names which correspond to 284 persons), to have changed little because the writer has simply copied them, it turns out that only 51% of these name-records have identical first names and family names in both books. (16) (fig.4)

fig.4

Variation of medieval persons names (about 1400)

N = 284 persons

Percent of persons with...

identical fam.-name identical first-name		fam.- <u>and</u> first-name
61 %	78 %	51 %
3 first char. fam.-name are identical 2 first char. first-name are identical		fam.- <u>and</u> first-name
88 %	94 %	83 %
3 first char. skeleton fam.-name are identical 2 first char. skeleton first-name are identical		fam.- <u>and</u> first-name
89 %	98 %	88 %

With the help of the identification code, 249 (88%) out of 284 name-records from the old "Bürgerbuch" are correctly linked to the corresponding name in the second book. An identification-code composed from the first three characters of the original family name and the first two characters of the original first name would have given us only 83% of correctly grouped names. The skeleton gives a further 5%. If we had only used the rules for skeleton processing, without the standardization of the entire names, (17) we probably would have achieved the same result. However, the name skeletons distinguish differences between the names more easily; the same code does not group too many names. An alternative code, composed of part of the original first and family names without processing, would have to be several positions longer. Using such a code, several name-records which should have been grouped are not; e.g. only 77% of the identical name-records would have been grouped if the family name had not gone through the skeleton processing, but had been increased to four instead of three letters. Thus the identification-code should create an ideal balance between the "faculty of grouping" and the "faculty of discriminating".

The 35 name-records (12%) from the first "Bürgerbuch" which cannot be correctly linked using the identification-code are made up of 12 name modifications (4%) and 23 written variations (8%) of names. (The latter are not taken into consideration by the skeleton program.) By changing the identification-codes of these 35 name-records manually (6% of the total number of the test name-records of the first and second "Bürgerbuch") (18) the researcher can group these records correctly.

If we examine the family name and first name separately, we find that only six first names (2%) are not correctly linked and must be corrected manually. This can be explained by the fact that family names vary six times more than first names.

In the "Bürgerbuch":

Out of 8400 family names 3706 are different. (19)

Out of 8000 first names 621 are different. (19)

As the first name is the original form of designation of a person it is quite constant and does not vary often. Therefore one might suppose that a sorting process with the skeleton of the first name as first sorting key would group a large part of the name-records correctly. (20) However, we have created our output list for identification on the basis of the family name code as first sorting key, for the following reasons: firstly, it is easier to identify persons if members of the same family are together in the file; secondly, several first names are very common and would create large unwieldy groups (the most frequent first names Johannes, Perrodus and Petrus appear in 24% of the name-records) and thirdly, several first names appear to be similar and it is difficult to say whether they are used indifferently (e.g. Petrus, Perrinus, Perrodus). The researcher, with his experience, can eliminate the incorrectly identified name-records by changing the identification-code or by adding a number in the sixth position of the identification-code in order to distinguish two identical name records which belong to different persons. In the above mentioned test, 6% of the identification-codes had to be modified. This 6% would represent 510 identification-codes out of the total of 8500 name-records in the whole file. (21) In the overall project, we changed the identification-code of 470 name-records, which brings us

fairly close to the 6% modified in the test. Half of these 470 name-records show a complete modification of the family name. (230 = 3% of all name-records) 1200 of the name-records have an additional number in the sixth position of the identification-code. (22) 670 (8%) name-records of the file have a question mark in the sixth position to show that they are unidentified.

Automatic test processes, such as the likelihood test (23) and other employed in studying demography in modern times; can hardly ever be used in studying medieval sources because of the lack of extra identifying characteristics in these sources. (24) In our project another element renders the identification even more difficult. In most other projects of record linkage, there are two or more distinct files which the researcher knows are related. (25) (e.g. There will be one reference in a parish register for the birth of a person and a corresponding one for his death in another register). If, as in our case, one identifies persons within the framework of one source, one cannot be sure how often the same person is mentioned. A systematic check of the validity of each identification is not possible as it is more difficult to judge whether two or more names should be grouped together than to discover a linkage between two separate files.

The use of electronic data processing in the identification process simplifies the work of the historical inquirer by creating an efficient identification-code and by organising and displaying all the data in a concise and clear form. It would be useful if a process existed which could group similar names regardless of different languages and regional particularities. Our identification-code could not fulfil the above requirements. Other researchers think that the Russel-Soundex-Code does not satisfy this universal role either. (26) The Soundex-Viewex-Code seems interesting, because it not only takes into consideration the soundex-code, which accounts for letters with similar sounds, but also letters which can appear similar in script form. (27) This code equates identical written letters, for example "c" with "t" two characters which are not easily distinguished in a manuscript. However, the rules of this code, as they were used in the Philadelphia Project, cannot be freely applied to other projects. The algorithm created by Guth seems useful. (28) In this procedure, a code is not created, but names are compared by employing the order and the identity of their letters. This operation is independent of the data and does not require standardization of the names or skeleton processing. It remains to see whether this algorithm can be employed with medieval data. (29)

One could also invent an identification code where the standardization of the names would be worked out manually step by step. One could imagine a file which would be sorted, using the first characters of the first and family name, and then would group similar name-records together. The successive standardization of names would require an extended sort-key, which, as the number of its positions increased, would make the identification-code more and more precise. This identification process would only apply to the standardization of the names and a corresponding code, to serve as a reference to the name-record in other files, would be created from the standardized first and family name. (30) The advantage of such a process would be that one could start the real identification step right at the beginning, without the skeleton program and without having to create a standard-name list. (31) However, 40% of the family names and 20% of the first

names would have to be modified by hand (32) and these corrections would require considerable time and work because the names are longer than the identification-code. There is another disadvantage when using this method: if the standardization is not produced by an automatic program, it would be arduous to introduce additional sources into the project. Nevertheless, one could imagine a standardization of the names by a program resembling the skeleton program which would be as follows: C/MAGENBERG/MACKENBERG/*etc.

Such a standardization program for our 8500 family names would require 2000 change commands, (33) which would be too laborious for a single project. A program for the first names, with about 300 change commands would however be feasible.

Another way to standardize names would be to use a philological approach. (34) In this case, the family and first names are reduced to their lemmas and then the different variants of lemmas are attributed to their corresponding lemma with the help of a lemma dictionary. This approach is especially interesting to someone studying philological and language questions. The historical inquirer who is interested in the background and life of a person will only use lemmatisation as a reference in his identification process. The relationship between the original written form and its lemma forms the philological basis for this above identification. Then it is up to the historian to decide whether names, which from the philological points of view are different, can be attributed to the same person. As it is the person and not the name itself which interests the historian, we chose a purely pragmatic way of building our identification-code which of course does not exclude an subsequent philological study of the names.

The identification-code is not only employed for the identification of name-records. It is also used as a reference key in all the relational files. For instance, it is the identification-code which refers to the person in the kinship relation or possessions file. It is possible to link to family names or to first names through the data base but it is an advantage if the identification-code shows - in a reduced form - the first and family names. Therefore the code should not be in numerical form or be so condensed that it becomes unintelligible.

Our aim in this paper is not to present an ideal solution to the problem of identifying persons in medieval sources. In theory it is only too easy to present many marvellous solutions which would only need to be put into practice in order to show their advantages and disadvantages. In this paper, however, we would like to emphasize some points which seem to be important when studying medieval data. As most of the identification surveys of the Middle Ages are based solely on name data, these names should be kept in their original form throughout the project. Any modification should be made on a second copy of each name. An identification-code, created from a reduced form of the names, should be used as a reference key and as basis for the identification. The most difficult methodological problem is to develop a process of optimal value which automatically recognize similar names as identical. It is also important to be able to correct and improve the identification of the persons when necessary and this requires an automatic processing run which create extra relational files from the data of the identified name-records. The essential role of the computer is to arrange and manage the data. It therefore facilitates the identification process which cannot be fully automated.

FOOTNOTES

1. Basic articles with bibliographical information concerning identification processing or record linkage: Wrigley, E.A., (ed.) Identifying People in the Past, London 1973. Concerning the Middle Ages: Herlihy, D., Problems of Record Linkages in Tuscan Fiscal Records of the Fifteenth Century, p. 41-56. An aid to see whether one should use automatic processing in identification or not: Winchester, I., Priorities for Record Linkage, A Theoretical and Practical Checklist, in: Clubb, J.M./Scheuch, E.K., (ed.) Historical Social Research, The Use of Historical and Process-Produced Data, Stuttgart 1980. Hershberg, T./Burstein, A./Dockhorn, R., Record Linkage, in: Historical Methods Newsletter, 9, 1975/76, p. 137-165. Winchester, I., The Linkage of Historical Records by Man and Computer: Techniques and Problems, in: Journal of Interdisciplinary History, 1, 1970, p. 107-124. Newcombe, H.B., The Design of Efficient System for Linking Records into Individual and Family Histories, in: American Journal of Human Genetics, 19, 1967, p. 335-359. Thaller, M., "Record Linkage" mit CLIO, in: Clio-Sysbull, 3, 1980. Apart from Herlihy, these authors do not deal exclusively with medieval identification problems. see note 24.
2. The "Freiburger Bürgerbuch (1341-1416)" project is subsidised by the Swiss National Fund. It is supervised by Prof. P. Rück, Univ. of Lausanne. The technical side is organised by H.M. Bächler, IBM Zürich. For the person-relations the data base system has been in operation since summer 1979. The QBE (Query by Example) data base system was chosen, using the relational concept. See Bächler, H.M./Rück, P., A Prosopographical and Topographical Evaluation of the First Bürgerbuch of Freiburg, Switzerland (1341-1416), in: Computers and the Humanities, 12, 1978, p. 137-144. Query-by-Example, Programm Description/Operations Manual, IBM, doc-nr: SH20-2077, and Query-by-Example, Terminal User's Guide, IBM 1978, doc-nr: SH20-2078-o.
3. Edition: Bonfils, Y./Vevey de, B., Le Premier Livre des Bourgeois de Fribourg (1341-1416), Fribourg 1941. A second book was started in 1416. All living "Bürgers" were transcribed into this new book. These records were taken into consideration for this project. Name-records total: 8536, inscriptions total 2609.
4. Hershberg (1975/76), p. 140 ss.
5. Other names given to the identification-code: derived sorting key, grouping variable. Normally, the "pocket variable" should only be considered as an identification-code after the identification process by the historian. In this article, we do not make this distinction.
6. Only the first name is put in the nominative form.
7. CMS-Editor is used. CMS (Conversational Monitor System) is a conversational system produced by IBM which runs on the VM (Virtual Machine) system. The editor of CMS is used for text input and output. Manuals: CMS-Users Guide, IBM, doc-nr.: GC20-1819, and IBM Virtual Machine Facility/370: CMS Command and Macro Reference, Release 3 PLC 1, IBM 1976, doc-nr. GC20-1818.

8. For further information: see Bächler/Rück (1978), p. 141 ss.
9. see note 4.
10. Hershberg (1975/76), p. 144: The pocket-variable should contain all possible identical units.
11. For the statistical results see p. 13.
12. The residence and personal possessions of a person are not included in the first stage of identification. In order to identify the above, it is necessary to identify the houses and this can only take place after a full identification of persons.
13. In figure 1 two inscriptions are represented: folio 23, recto, 1st inscription, and folio 23, recto, 2nd inscription, see Bonfils/Vevey de (1941), p. 45.
14. The relatives who are mentioned within an inscription have been coded in a two dimensional table, which does not appear in fig. 1. This table is as follows for the 1st inscription:

	1.Pers.	2.Pers.	3.Pers.	4.Pers.
1. Person		30		(Code 30=son of)
2. Person	10			(Code 10=father of)
3. Person				
4. Person				

see Bächler/Rück (1978), p. 138 ss.

15. See Herlihy in: Wrigley (1973), p. 50.
16. For their 19th century data, Kelly found between 64% and 84% of linked records with identically spelt surnames.
17. 108 cases of standardisation of complete family names and 17 of complete first names (see above).
18. The identification-code is only necessary in one of the two name-records.
19. Spelling variations are counted.
20. The first-name file only serves as an aid in the identification processing.
21. The chosen test-data represent 7% of all the data. In the complete file, it is possible to have several linkages.
22. The highest number of entities grouped together under one identification-code is 8. In most cases this number is only two or three.
23. For the different weighting systems see articles in: Wrigley (1973), Hershberg (1975/76), p. 144 ss., Newcombe (1967), Winchester (1970).
24. We do not know of a medieval historian who uses an automated comparison-step. Most medieval researchers use a system of files to facilitate identification, cf. Batori, I., Sozioökonomische Untersuchungen in süddeutschen Städten des 15. und 16. Jh., Programmabläufe-Erfahrungen- Er-

gebnisse, in: Irsigler, F., (ed.) *Quantitative Methoden in der Wirtschafts- und Sozialgeschichte der Vorneuzeit*, Stuttgart 1978, p. 24-42, (28). Lipkin, B.S./Lipkin, J., *Data Base Development and Analysis for the Social Historian: The Educational Status of the Beneficed Clergy of the Diocese of Hereford, 1289-1539*, in: (Gilmore-Bryson, A., (ed.) *Medieval Studies and the Computer, A Special Issue of: Computers and the Humanities*, 12, 1978, p.113-125, (p. 118 ss.): The Information in the files, which are sorted by the names, is assembled by hand in a new file. Identifications are controlled by references. A more efficient method is being developed. see also: *Kommentiertes Parallelregister und Untersuchungen*, vol. 2.1, in: Schmid, K., (ed.) *Die Klostergemeinschaft von Fulda im frühen Mittelalter*, 3 vol., Münstersche Mittelalterschriften, 8, 1978: The computer assembles the documents in "Kommentierten Parallelregistern" and thus creates an instrument of research. See also the different articles in: Fossier, L./Vauchez, A./Violante, C., *Informatique et Histoire Médiévale*, Rome 1977. Even in research on the 19th century, one encounters difficulty in establishing criteria for computerised comparison. Hershberg (1976), p. 158.

25. Hershberg (1976), p. 139: On the relationship between data files.
26. Winchester in: Wrigley (1973), p. 130. Hershberg (1975/76), p. 141. The rules of the Russel-Soundex-Code in: Winchester (1970), p. 115.
27. Coding rules for this code in: Hershberg (1975/76), p. 142 s.
28. Guth, G.J., Surname Spellings and Computerized Record Linkage, in: *Historical Methods Newsletter*, 10, 1976, p. 10-19.
29. On different coding rules see Wrigley in: Wrigley (1973), p. 98 ss. Winchester (in: Wrigley (1973), p. 131) gives a bibliographical survey of different methods.
30. F.i. a code made up of the first five letters of the familyname and of the first three of the first name.
31. Identification by standardization of the names in: Herlihy, D./Klapisch, C., *Les Toscans et leurs familles, Une étude du catasto florentin de 1427*, Paris 1978, p. 101 ss., and Herlihy (1973) Luzzati, M./Sprugnoli, R./Carratori, L., *The Use of Computers in Editing Medieval Baptismal Records in Pisa (1457-1509)*, in: *Computers and the Humanities*, 12, 1978, p. 155-164, (p. 157). While Herlihy retains all the variations of a name, Luzzati, on the other hand, keeps them in a special file. See also Metman, J./Auzary, B., *Les archives judiciaires et leur traitement informatique*, p. 361-377, and Langlois, M., *Contrôle et enrichissement d'un instrument de travail ancien, Le fichier des noms des parties en procès devant le parlement de Paris*, p. 381-392, in: *Informatique et Histoire médiévale* (1977),

32. see fig. 3.
33. Such a normalization could not be carried out with a CMS-procedure, because of the numerous calculations involved. Gervers has suggested an index which would give a standard form of every given variant. Gervers, M., *Medieval Charters and the Computer: An Analysis Using Mark IV*, in: *Computers and the Humanities*, 12, 1978, p. 127-136, (p. 131).
34. One can refer principally to the work of the "Sonderforschungsbereich Mittelalter: Personen und Personengruppen" in Münster, who publish an annual report with a bibliography in the "Frühmittelalterliche Studien". This work concentrates on the early Middle Ages, where only first names were used. The latest article on this theme are: Kamp, H., *Ein Algorithmus zur automatischen Lemmatisierung von Personennamen*, and Geuenich, D., *Die Lemmatisierung und philologische Bearbeitung des Personennamenmaterials*, in: Schmid (1978). Other institutes working on similar subjects are: in Florence: "Centro per lo studio delle civiltà barbariche in Italia". (cf. Arcamone, M.G., *Per lo spoglio elettronico dell' onomastica Italiana Alto-medioevale*, in: *Informatique et Histoire médiévale* (1977), p. 193-195.) and, in Paris: "Institut Historique Allemand de Paris" (*Prosopographia regnorum orbis latini*, PROL). (cf. Werner, K.F., *Die Forschungsbereiche des Deutschen Historischen Instituts in Paris, ihre Schwerpunkte und Projekte*, in: *Francia*, 4, 1976, p. 722-748.) see also annual report in "Francia".
35. Like f. i. the Russel-Soundex-Code.